Progress and Prospects of the Human-Robot Collaboration
Arash Ajoudani, Andrea Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, Oussama Khatib

To cite this version:

HAL Id: hal-01643655
https://hal.archives-ouvertes.fr/hal-01643655
Submitted on 21 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Progress and Prospects of the Human-Robot Collaboration

Arash Ajoudani, Andrea Maria Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, and Oussama Khatib

Received: date / Accepted: date

Abstract Recent technological advances in hardware design of the robotic platforms enabled the implementation of various control modalities for improved interactions with humans and unstructured environments. An important application area for the integration of robots with such advanced interaction capabilities is human-robot collaboration. This aspect represents high socio-economic impacts and maintains the sense of purpose of the involved people, as the robots do not completely replace the humans from the work process. The research community’s recent surge of interest in this area has been devoted to the implementation of various methodologies to achieve intuitive and seamless human-robot-environment interactions by incorporating the collaborative partners’ superior capabilities, e.g. human’s cognitive and robot’s physical power generation capacity. In fact, the main purpose of this paper is to review the state-of-the-art on intermediate human-robot interfaces (bi-directional), robot control modalities, system stability, benchmarking and relevant use cases, and to extend views on the required future developments in the realm of human-robot collaboration.

1 Introduction

The fast growing demands for service robot applications in home or industrial workspaces have led to the development of several well-performing robots equipped with rich proprioception sensing and actuation control. Such systems that range from robotic manipulators [Albu-Schäffer et al., 2007] to full humanoids [Tsagarakis et al., 2016, Ott et al., 2006, Kaneko et al., 2008, Radford et al., 2015] are expected to help the human user in various tasks, some of which require collaborative effort for a safe\(^1\), successful, and time and energy efficient execution. In fact, the integration of robotic systems in collaborative scenarios has seen an extensive and fast-growing research effort, a tentative estimation of which is provided in Fig. 1 by referring to the number of publications on this topic over the last two decades.

Physical human-robot collaboration (PHRC), which falls within the general scope of physical human-robot interaction (see [De Santis et al., 2008, Murphy, 2004, Alami et al., 2006]), is defined when human(s), robot(s) and the environment come to contact with each other and form a tightly coupled dynamical system to accomplish a task [Bauer et al., 2008, Krüger et al., 2009]. Ideally, each active component of such a system must be capable of observing and estimating the counterparts’ contributions to the overall system’s response through the fusion and processing of the sensory information [Argall and Billard, 2010, Ebert and Henrich, 2002, Lallée et al., 2012]. As a consequence, an appropriate reactive behaviour can be replicated (e.g. by the human from a set of obtained skills in previous attempts of performing a similar task) or developed to complement and improve the performance of the collaborative partners.

\(^1\) The problem of safety in human-robot interaction (HRI) and the related open issues have been extensively discussed in literature [Haddadin et al., 2009, De Santis et al., 2008, Alami et al., 2006]. Hence, our focus in this review paper will be on other important aspects of physical human-robot collaboration (PHRC).
Similar to the humans’ anticipatory (feed-forward [Shadmehr and Mussa-Ivaldi, 1994b]) or feed-back [Todorov and Jordan, 2002] mechanisms to develop a suitable motor behaviour, the collaborative robots’ response to sensory input can be achieved through model/knowledge based techniques [Tamei and Shibata, 2011, Ogata et al., 2003, Kimura et al., 1999, Magnanimo et al., 2014], the implementation of feedback controllers with pre-set interaction modalities [Petersen et al., 2016c, Donner and Buss, 2016a] or a combined approach [Rozo et al., 2013, Petersen et al., 2016b, Lawitzky et al., 2012b, Palunko et al., 2014]. A key strategy in this direction is the establishment of a shared authority framework in which the significant capabilities of both humans and robots can be exploited. For instance, humans’ significant cognitive abilities in learning and adaptation to various tasks demands and disturbances can be used to supervise the collaborative robots’ superior physical capabilities. The increasing slope in the research community’s interest towards the integration of human intention in real-time adaptation of the robot behaviour is illustrated in Fig.1.

With this in mind, the main purpose of this paper is to review the state-of-the-art on the key enabling technologies to achieve a seamless and intuitive human-robot collaboration. Although hardware-related components are among the most critical to achieve this goal, a remarkable effort has been recently devoted to overview the underlying progress in terms of communication [Wang et al., 2005], sensing [Tegin and Wikander, 2005], and actuation [Vanderborght et al., 2013a] developments. Hence, our focus in this review will be on other relevant key elements of the HRC, i.e. human-robot interfaces (bi-directional), robot control modalities, system stability, benchmarking and relevant use cases. In addition, with the aim to achieve a reasonable convergence of views for the required future developments, our focus will be mostly on the physical aspects of the human-robot collaboration.

2 Interfaces for Improved Robot Perception

Humans embrace a diversity of experiences from working together in pairs or groups. This has contributed to the development of implicit and explicit communication standards so that task-related information can be perceived and communicated intuitively [Sebanz et al., 2006]. In fact, one of the main objectives in the realm of physical human-robot collaboration is to design and establish similar communication standards so that the robot is aware of human intentions and needs in various phases of the collaborative task [Klingspor et al., 1997, Bauer et al., 2008]. Despite the fact that the robotic replication of the human sensory system is hardly possible with the current technology, the understanding and implementation of the underlying communication principles can potentially lead to an enhanced physical human-robot interaction performance [Reed and Peshkin, 2008].

A widely known example of such a communication interface is built on the use of visual [Perzanowski et al., 1998, Agravante et al., 2014a, Morel et al., 1998] or language commands [Medina et al., 2012a, Miyake and Shimizu, 1994, Petit et al., 2013], as user-friendly means of communicating with a robot, from the human standpoint, given that the human is not required to learn additional tools. The use of head, body or arm gestures are common examples in the areas of human-robot interaction and collaboration [Li et al., 2005, Carlson and Demiris, 2012]. In this direction a method to interpret the human intention from the latest history of the gaze movements and to generate an appropriate reactive response in a collaborative setup was proposed in [Sakita et al., 2004]. Authors in [Hawkins et al., 2013] developed a vision-based interface to predict in a probabilistic manner when the human will perform different subtasks that may require robot assistance. The developed technique allows for the tracking of the human variability, environmental constraints, and task structure to accurately analyse the timings of the human partner’s actions.

Although such interfaces appear natural to the humans, their usage is mostly limited to activating high-level robot operations and the task complexity can potentially prevent the robot from deriving the desired sensorimotor behaviour from these higher-level modalities. In fact, a large degree of robot autonomy, which is far beyond current capabilities of the autonomous robots, is required for vision or auditory based interfaces to function on a wide range of applications.

An alternative approach to the design of human-robot interfaces recognises the use of force/pressure sensors in contact to anticipate the objective of the human partner and/or to control the cooperation effort. Due to the simplicity of the underlying mechanism, it has been explored in several applications, examples include collaborative object transportation [Ikeura and Inooka, 1995a, Kosuge and Kazamura, 1997a, Al-Jarrah and Zheng, 1997a, Tsumugiwa et al., 2002a, Duchaine and Gosselin, 2007, Agravante et al., 2014a, Gribovskaya et al., 2011a, Rozo et al., 2014, Rozo et al., 2015, Adams et al., 1996], object lifting [Evrad and Kheddar, 2009, Evrad et al., 2009], object placing [Tsumugiwa et al., 2002a, Gams et al., 2014], object swinging [Donner and Buss, 2016a, Palunko et al., 2014], posture assistance [Ikemoto et al., 2012, Petersen and Babic, 2013], and industrial complex assembly processes [Kruger et al., 2009] (see also Fig. 2).

In most of the above techniques, the interaction forces/torques are used to regulate the robot control parameters and trajectories following the admittance [Duchaine and Gosselin, 2009, Lecours et al., 2012] or impedance [Tsumugiwa et al., 2002a, Agravante et al., 2014a] causality [Hogan, 1985]. Notwith-
standing the wide margin of applications, collaborative tasks that involve simultaneous interaction with rough or uncertain environments (e.g. co-manipulative tool-use) can induce various unpredictable force components to the sensor readings [Peternel et al., 2014]. This can significantly reduce the suitability of such an interface in more complex interaction scenarios since it can be difficult to distinguish the components related to the active counterpart(s) behaviour from the ones generated from the interaction with the environment.

Bio-signals such as electromyography (EMG) and electroencephalography (EEG), or other physiological indices such as electrodessinal activity [Pecchinenda, 1996, Rani et al., 2006] can be used to anticipate the human intention in PHRC. In particular, due to the adaptability and ease-of-use of EMG measurements, they have found a wide range of applications in human-in-the-loop robot control such as: prostheses [Farri et al., 1996, Jiang et al., 2009, Farina et al., 2014, Castellini et al., 2014, Strazzulla et al., 2017], exoskeletons [Rosen et al., 2001, Fleischer and Hommel, 2008] and industrial manipulator control [Vogel et al., 2011, Peternel et al., 2014, Ajoudani, 2016, Gijsberts et al., 2014]. Peternel et al., used EMG signals to anticipate the stiffening/complying behaviour of a torque controlled robotic arm in a co-manipulation task [Peternel et al., 2016c]. Through this interface, the leading/coming roles of human and the robot counterpart were estimated online. In another study, Bell et al., used EEG signals to command a partially autonomous humanoid robot through high-level descriptions of the task [Bell et al., 2008].

A remarkable use of bio-signals in the development of HR interfaces is to estimate the human physical (e.g. fatigue) or cognitive (anxiety, in-attention, etc.) state variations that might deteriorate the collaborative robot’s performance. The authors in [Rani et al., 2004] developed a method to detect human anxiety in a collaborative setup by extracting features from EMG, electrocardiography (ECG) and Electrodermal responses. In a similar work, the human physical fatigue was detected and used to increase the robot’s contribution to the task execution [Peternel et al., 2016b].

Although an interface that is build on a unique source of sensory data can configure a pre-defined robot behaviour in collaborative settings, the underlying functionality is limited and cannot be easily generalised to cross domain scenarios. For instance, the use of visual feedback for the estimation of the exchanged amount of energy between the counterparts is less effective than the use of force or pressure sensors. Similarly, the use of bio-signals such as EMGs for tracking of the human limb movements may result in less accurate performances in comparison to the external optical or IMU-based (inertial measurement unit) tracking systems [Corrales et al., 2008]. To address this, a combined approach, associating multi-modal sensory information to different robot control modalities (commonly known as multi-modal interfaces [Mittendorfer et al., 2015, Peternel et al., 2016c]), can be exploited. In this direction, the authors in [Agravante et al., 2014a] proposed a hybrid approach by merging vision and force sensing, to decouple high- and low-level interaction components in a joint transportation task where a human and humanoid robot carry a table with a freely moving ball on top (see also [Rozo et al., 2016]). A similar work proposed a multi-modal scheme for intelligent and natural human-robot interaction [Bohme et al., 2003] by merging vision-based techniques for user localisation, person localisation and person tracking and their embodiment into a multi-modal overall interaction schema.

In a similar fashion, voice commands were used to pause, stop or resume the execution of a dynamic co-manipulation task, the control parameters of which regulated by an EMG based interface (See Fig. 3). In this work, an external tracking system detected the human arm configuration to regulate the robot task frame in real-time. By the same token, authors in [Yang et al., 2016] developed a multi-modal teaching interface on a dual-arm robotic platform. The interface was built on the use of EMG on the user arm and force sensors on robot end-effector. In this setup, one robotic arm is
connected to the tutee’s arm providing guidance through a variable stiffness control approach, and the other to the tutor to capture the motion and to feedback the tutees performance in a haptic manner. The reference stiffness for the tutors arm stiffness was estimated in real-time and replicated by the tutee’s robotic arm. Ivaldi et al., [Ivaldi et al., 2016] studied multi-modal communication of people interacting physically with the humanoid iCub to build objects (see Fig. 4). Participants would naturally gaze at the robot’s hands or face to communicate the focus of attention of the collaborative action, while speaking to the robot to describe each action. The authors found that individual factors of the participants influence the production of referential cues, both in speech and gaze: particularly, people with negative attitude towards robots avoid gazing at the robot, while extroverted people speak more to the robot during the collaboration. The robot, controlled in impedance with a low stiffness at joints, switched to zero-torque control when the humans were grasping the robot arms covered by a tactile skin (enabling precise contact estimation [Fumagalli et al., 2012]) and giving the voice command to “be compliant”. This multi-modal command strategy allowed the participants, not experts in robotics and mostly interacting with a robot for the first time, to physically move the robot arms in an intuitive way, without generating anxiety for their safety as reported by the participants in their interviews. Despite the lack of experience, all the human participants were able to interact physically with the robot to teach the task. Facilitated probably by the child-like appearance of the iCub, the participants naturally acted as teachers/caregivers, in line with the observations of [Nagai and Rohlffing, 2009]; however, it has to be remarked that in that situation of physical interaction the authors did not observe exaggerated movements typical of parental motionese/scaffolding situations in HRI, where often there is little physical interaction and more social cues as gaze or gestures. These results are in accord with previous work presented, see e.g. [Kilner et al., 2003, Ugur et al., 2015].

The improved performance of the multi-modal interfaces in the generation of compound robot behaviours that are required to execute more complex collaborative tasks has shifted the attention towards the usage and fusion of multi-source sensory information. Results of Google Scholar suggest that over 76% of the publications in the area of human-robot collaboration used multi-modal interfaces in year 2015. Nevertheless, the inclusion of more communication channels in the development of the intermediate interfaces will potentially contribute to an increase in the human cognitive burden and the low level robot control complexity. This may affect the intuitiveness of the interface and result in an excessive human effort to operate a specific robot modality. A solution to this issue can be obtained by the introduction of shared communication modalities [Green et al., 2008, Lackey et al., 2011, Lackey et al., 2011]. Alternatively, robotic learning techniques such as: gradual mutual adaptation [Ikemoto et al., 2012, Petermel et al., 2016a], reinforcement learning [Palunko et al., 2014] or learning from demonstration [Evrard et al., 2009, Lawitzky et al., 2012a, Rozo et al., 2015] can be exploited to weaken the communication loops’ demands (e.g. bandwidth, number of feedback modalities) due to an increased level of robot autonomy.

3 Interfaces for Improved Human Perception

The visual and auditory systems of the humans provide powerful sensory inputs that contribute to a fast and accurate perception of the movement kinematics and the environment, and a constant update of the internal models. The role of such sensory inputs in dynamic perception of the environment, e.g. anticipating the weight of an object through vision [Gordon et al., 1993], and estimating a required amount of force to move the object along a pre-defined path [Johansson, 1998], has also been investigated.

During collaboration and dyadic interaction, mutual gaze and joint attention are common ways of conveying information [Tomasello, 2009]. Such mechanisms are often implemented in robots to make the interaction more effective by providing additional back-channels. For example in [Ivaldi et al., 2014] the robot was equipped with anticipatory gaze mechanisms and proactive behaviours, increasing the pace of the interaction and reducing the reaction time of the human to the robot’s cues.

In a similar study, Dumora et al. [Dumora et al., 2012], studying haptic communication between an operator and a robot for a bar transportation task, observed that wrench measurements provide incomplete information to detect the operator’s intent of motion. This has been also observed in [Reed, 2012] for cooperating dyads able to communicate
only through physical interaction\(^3\). Although haptic feedback can temporarily be sufficient to achieve sub-tasks, if contextual information about the task phase are provided, for the interaction to be efficient more non-verbal cues are necessary to recognize the partner intent during collaborative tasks and synchronize the dyadic activity. In this sense, joint attention or proactivity can improve the mutual awareness, hence the task performance, as shown by [Ivaldi et al., 2014] for a dyadic learning task.

An enhanced perception of the environment through human visual feedback can also be achieved by Augmented Reality (AR) approach, enabling the human partner to observe and review a plan with the robot prior to execution [Green et al., 2009, Glassmire et al., 2004] (see Fig. 5). Head Mounted Displays, that were mostly used in research labs or in very specific, critical applications (e.g. aero-space and military), are nowadays becoming a consumer product: the Rift by Oculus and the Gear VR by Samsung (among others), are the first modern and large scale examples of a market that is on the verge of the explosion. Potential drawbacks of the AR based approaches can be the information overloading, limited privacy (e.g. augmenting without permission) and additional cost, that may limit the expected performance of such systems in collaborative settings.

Haptic information provided by receptors in human limbs (fingertips, arm skin, etc.), on the other hand, represents a very important and complementary input to explore the external environment and for everyday task accomplishments. Since in most collaborative scenarios the human partner comes to physical contact(s) with the object and/or the robot in a closed dynamic chain, a large amount of meaningful information can be perceived by the human receptors. Nevertheless, as regards specific interaction scenarios in which the human perception of the environment is influenced by the task or environmental conditions, the use of artificial sensory systems in the provision of haptic feedback to the human partner can be beneficial.

Although replicating the richness of human sensory apparatus through artificial systems remains a daunting task, the state-of-art reviews several non-invasive techniques to present haptic stimuli to robot operators, by delivering different types of stimuli to the human limb. This is not only to provide cheaper and easily applicable feedback systems can replace full force feedback with none or small performance reduction, but also to resolve fundamental issues such as stability in closed loop [Tegin and Wikander, 2005]. In addition, such sensory substitution techniques can also be used to provide the human with feedback about the advanced robotic perception of the environment (e.g., depth sensors [Plagemann et al., 2010]), that may not be covered by the human sensory system. Examples include vibrotactile, electrocutaneous or mechanical pressure [Shannon, 1976, Nohama et al., 1995, Wang et al., 1995], to convey information about force, proprioception, and/or texture. By the same token, previous works recognized different types of non-invasive sensory substitution techniques, roughly dividing haptic information into low-frequency, e.g. force related and high-frequency [Ajoudani et al., 2014], e.g. acceleration related [Godfrey et al., 2013]. In a similar vein, a haptic force feedback interface was developed in [Yang et al., 2016] to provide a tutor with information about the deviations of the tutee arm from desired trajectories in a human-robot collaboration setup that was developed as a teaching interface through a dual-arm robotic platform. The amount of force was related to the kinematic error between the desired and the tutee’s replicated trajectories in real-time.

Table 1 presents an overview of the interfaces used in literature with the purpose of enhancing the robot and human perception of the task and the environment.

### 4 Interaction Modalities

This section aims at presenting an overview of the different strategies to endow the robot with interaction capabilities. While the interfaces and the underlying perception mechanisms are dealt with in Section 2, this part discusses different approaches to link the perception inputs to their actual effects in terms of robot behaviour.

As mentioned earlier, a very important aspect in setting up a successful PHRC framework is to achieve robot adaptivity to the task conditions and environmental constraints. One of the pioneering works in this sense are due to Hogan [Hogan, 1985], presenting the well-known impedance (and admittance, as a dual representation) control framework. Although the original work did not cover explicitly the case of human-robot interaction, the same formalism has been further developed in [Kosuge and Kazamura, 1997b, Tsuamugiwa et al., 2002b, Albu-Schäffer et al., 2007, Albu-Schäffer et al., 2010].

Fig. 5: An example of providing haptic and visual feedback (via AR) to the human counterpart in collaborative settings [Glassmire et al., 2004] (image courtesy of M. O’Malley).

---

\(^3\) The effect of human sensorimotor learning is very crucial in such scenarios as the perception of the environment and the task by two experts will be different from two naive operators.
2007, Gribovskaya et al., 2011b] to explicitly account for the human as the source of interaction forces. In particular, Kosuge et al. proposed several control strategies to allow the human and a robotic manipulator to co-manipulate heavy and bulky objects in [Kosuge and Kazamura, 1997b]. In this scenario, that is quite common in the literature, the robot is responsible for carrying a payload, while the human steers the object to a desired position or along a desired path.

A similar concept was investigated in [Tsumugiwa et al., 2002b], where the authors augmented the sensor space with virtual forces to ease the human to complete a collaborative peg-in-hole task. The work in [Bestick et al., 2015] presented a framework for the parameter and state estimation of personalised human kinematic models using motion capture data. Data-driven musculoskeletal models are incorporated into the robot interaction controllers to account for differences of individuals and to provide personalised interaction/assistance capabilities. The authors in [Gribovskaya et al., 2011b] proposed an adaptation scheme for impedance parameters to enhance robot learning capabilities. In particular, Gaussian Mixture Models (GMM) were used to learn the task model, thus allowing the robot to anticipate the partner’s intentions and to adapt its motion according to perceived interaction forces. The same concept was extended in [Kronander and Billard, 2014]. Similarly, the work in [Ragaglia et al., 2016] proposed a kinaesthetic teaching strategy (see also [Lee and Ott, 2011]), based on impedance control, which is able to achieve both high accuracy and repeatability of the motion of the robot, as it is hand-guided by the human operator. In [Mörfl et al., 2012], the authors proposed dynamic role exchange mechanisms and applied the approach to a collaborative transportation task. Here, the impedance control concept is used to simultaneously account for the load sharing problem, the motion control of the manipulated object and the minimisation of internal wrenches.

Alternative techniques aim at developing implicit interaction capabilities through real-time motion generation and trajectory re-planning methods. Visual perception is one of the commonly adopted sensory input in this area due to the possibility of tracking the human motion in real-time. In the work of Ebert et al. [Ebert and Henrich, 2002], a set of stationary 2D ceiling cameras were used to predict possible unsafe situations within a collaborative industrial work-cell. The system was able to predict possible collisions and to promptly command a stopping trajectory to the manipulator. Within a similar setting, in [Bascetta et al., 2011] the system is trained to predict long-term walking trajectories of the human worker in order to proactively trigger an appropriate interaction behaviour. The approach used GMM for modelling the data set of acquired trajectories and a Hidden Markov Model (HMM) for the online prediction phase.

A relatively recent line of research in the realm of human-robot collaboration is to generate natural human-like robot motions for an enhanced acceptability and compatibility to the human workspaces [Zanchettin et al., 2013b]. The applications of robotic learning techniques in this context has become indispensable. For instance, the authors in [Kim et al., 2006] exploited a motion capture database filled with human movements to generate human-like motion primitives. In a similar vein, authors in [Khatib et al., 2009] developed a musculoskeletal model to analytically describe the natural human motion and to minimise human fatigue. Calinon et al. [Calinon et al., 2010] proposed a probabilistic learning strategy to reproduce human-like motions through imitation. The approach is based on a combination of HMMs and Gaussian Mixture Regression (GMR) to learn and reproduce from a demonstrated set of data.

Given the variety of interfaces already discussed in Section 2 and the availability of different control strategies analysed so far, a natural field of new investigations is represented by multi-modal interaction: several modes among those discussed earlier in this Section are combined to form the interaction behaviour of the robotic manipulator. With the aim of mimicking the way humans interact with each other, the first combination of perception and control algorithms exploited the two mostly important senses: vision and touch (tactile perception). The first results on this topic were reported in [De Santis et al., 2007] where a control scheme, mainly targeting the enhancement of safety, has been proposed to account for both force and vision information. The authors in [Magrini et al., 2014] used visual information to understand the contact points between the human and the robotic manipulator. Another control architecture considering both visual and force information was presented in [Zanchettin and Rocco, 2015], where visual information was used to track an object while the human and the robotic manipulator are interacting by means of an impedance control scheme. In [Cherubini et al., 2013], the authors present a control scheme dealing with both visual perception and force information allowing the robot to infer human intentions during a collaborative screwing task. The different interaction modalities are sequenced based on the gathered sensor information. A multi-modal control strategy for a peer to peer collaborative carrying task has been presented in [Agravante et al., 2014b]. Here, the force and visual information are merged together to allow both safe and effective interaction. In particular, the former is mainly used to establish safety, while the latter is used to control the orientation of the transported object.

A more recent work in this area proposed a multi-modal interaction modality by implementing a hybrid impedance/force control in different Cartesian axes of the collaborative robot [Peternel et al., 2016c]. The force controller’s objective was to establish and maintain the contact between a sawing tool and the environment. Instead, the impedance controller operated on a combined EMG-visual tracking interface to
regulate the task stiffness and its frame in various phases of the collaborative task.

In general terms, the improved adaptability of the robotic platforms to unpredicted events has paved the way for the deployment of well-performing robots in complex interaction scenarios. However, the understanding the required levels of adaptivity from the two counterparts is still at its beginning, and a principled formalism is required to address the mutual human-robot adaptation requirements [Ikemoto et al., 2012].

5 Stability and Transparency of the PHRC Systems

When humans interact physically with robots, the robot control faces critical challenges in achieving performances while ensuring stability. Indeed, as well explained in [Buerger and Hogan, 2007], the stability of the human-robot system depends on the coupled dynamics. Interactions can alter both performance and stability of the two systems when coupled; hence, even if the two systems are stable in isolation, the coupled system may be unstable and not satisfy performance objectives. For example, a robotic manipulator may be very fast and precise in following a desired path along a known surface; a human may physically grab the robot to adapt the nominal trajectory to variability in the task: this interaction should always be “stable” regardless of the forces exerted by the human, gripping force, hand stiffness while grabbing and steering the robot, etc. Most controllers for such interacting systems regulate forces or dynamics behaviour of the robot at the interaction port (the location of the interaction, e.g., the end-effector). As mentioned before, a traditional control concept for achieving these performances is impedance control [Hogan, 1985], which consists in controlling the dynamic behaviour of the robot under the action of an external force, modelling the system as a spring-mass duo, with desired stiffness and damping.

Another desirable property of the robot control during interaction should be transparency [Lamy et al., 2009], e.g., in some applications, the robot should not oppose resistance to the force applied by the human operator when driving the robot, but it should rather move according to the intention of the human operator. This issue is particularly critical during co-manipulation, where an object is simultaneously held by the human and the robot. A similar performance is required in medical applications, such as assisted gesture guidance, but also in industrial contexts, such as assisted assembly.

Transparency is usually obtained by force feedback control. However, even transparency requires stability: a large bandwidth with high-force feedback gains is necessary to achieve a good transparency, but it has been shown in [Eppinger and Seering, 1987] that this condition may lead to instability if the environment is stiff (see in [Colgate and Hogan, 1989] the concept of “contact instability”), or more generally if its impedance is unknown or varying.

Stability in interaction control is therefore linked to the control of mechanical impedance. In traditional robots, driven by stiff DC actuators, impedance is obtained by active force control: while many recent robots equipped with pervasive joint torque sensing and controlled with fast low-level loops (e.g., the KUKA iiwa) can achieve good response in safe impedance ranges, interaction control can be more challenging in other robots with proximal sensing (such as the iCub [Fumagalli et al., 2012]). Indeed, time-delays in computing the active control and non-collocation (e.g., distance between sensors and actuation) can cause potential threats to the stability, reducing significantly for example the range of stable impedance values for interaction [Berret et al., 2011]. To increase the intrinsic compliance of the robot, there have been several robots equipped with elastic or flexible joints, and more recently with variable impedance actuators [Vanderborght et al., 2013b], which enable fast regulation of mechanical impedance at the joint level.

As well documented in [Buerger and Hogan, 2007], early works in interaction control assumed that impedance could be sculpted to achieve desired performances. However, achieving truly arbitrary impedance on any real robotic system is not possible, especially during interaction with humans, because the environment stiffness is unknown or varying, which may lead to coupled or contact instability [Colgate and Hogan, 1989]. This issue has been investigated using the concept of passivity: in [Colgate and Hogan, 1988], Colgate showed that when two stable systems with passive port functions are physically coupled, stability of the coupled system is guaranteed. However, passivity leads to very conservative gain tuning and constrain the apparent endpoint inertia to remain 50% higher than its real value, which prevents many co-manipulation applications with heavy robots [Lamy et al., 2009]. It is important to note here that in robots with joint torque sensing, the torque feedback can contribute to a substantial reduction of total inertia, while staying in a passivity framework [Albu-Schäffer et al., 2007].

Several studies showed that passivity may be too conservative and should be relaxed to increase performance during physical collaboration [Buerger and Hogan, 2006, Duchaine and Gosselin, 2008]. To relax the passivity constraint, Newman introduced the natural admittance control (NAC) [Newman, 1992], focused at reducing the apparent friction independent from inertia. Ficuciello et al. in [Ficuciello et al., 2014] proposed to use the robot’s redundancy to ensure a decoupled apparent inertia at the end-effector thus increasing the range of stable impedance parameters. Buerger & Hogan introduced the concept of “complementary stability” [Buerger and Hogan, 2007], which transforms the stability problem into a robust stability problem: by making the hypothesis that the human arm dynamics is unknown but...
reasonably bounded, the goal is to find the optimal tuning for a parameterized controller that is stable for the range of different environment dynamics. NAC and complementary stability were combined in [Lacevic and Rocco, 2011], but, as observed in [Dimeas and Aspragathos, 2016], their conservative gains were still hindering a smooth collaboration without efforts for the human operator. In [Dimeas and Aspragathos, 2016], the authors proposed online adaptation of an admittance controller gains on the basis of an instability index, computed in the frequency domain, that detects unstable behaviours of the human-robot coupled system by monitoring high-frequency oscillations in the force signal.

The main advantage of using admittance control schemes or optimal control schemes for adapting the gains and regulating the exchanged forces is that control theory provides rigorous frameworks and methods to prove the stability of the system and designing controls that are robust to perturbations. For interaction policies acquired or refined via machine learning techniques, such proofs are more difficult. If, for example, an interaction policy is learned from demonstration and refined through reinforcement learning (which is a common case in most robotics learning scenarios [Amor et al., 2009]), since the policy or its model are acquired on-line, and not known a priori, it is very difficult to provide stability proofs if not some conditions on the control bounds. In [Khansari-Zadeh and Billard, 2011], Khansari-Zadeh and Billard proposed a method, called SEDS, to learn stable parameterisations of dynamical systems used to reproduce robot movements learned by demonstrations. The method, exploiting Gaussian Mixture Models, ensures time-invariance and appropriate reaction to perturbations thanks to a proof of global asymptotic stability at the target. However their approach was focused on stabilising movement trajectories, not specifically to stabilising impedance during interaction. In [Khansari-Zadeh et al., 2014], Khansari-Zadeh et al. proposed an extension of their framework that combines the problem of learning motion trajectories and regulating the impedance during interaction, ensuring global stability. They provided sufficient conditions to ensure global asymptotic stability for movements in free-space, and passivity during persistent contact with a passive environment.

Stability and transparency are primordial to achieve a safe physical interaction, particularly to allow the robot to follow human guidance. These properties are fundamental for the robot to optimize the collaboration introducing other features such as legibility (i.e., the property of generating legible motions that can be easily understood by the human partners) and anticipation (i.e., the property of predicting the human intention, the goal of the collaboration, and optimizing the robot control policy to take into account the human
action, reducing the human effort or improving some shared performance criteria) [Sulp et al., 2015, Dragan et al., 2013]. Communication of intention in human-robot interaction increases with mutual practice, it is related to the arm impedance, to the continuous role adaptation between the partners [Mojtahedi et al., 2017] and it is also related to transparency [Jarrasse et al., 2008]: the same evidence holds for human-robot physical interaction [Dragan and Srinivasa, 2012, Jarrasse et al., 2013]. Transparency and stability are also crucial for optimizing the control in a way to provide assistance to the human while performing the task, as it is frequently done in robotic rehabilitation [Morasso et al., 2007].

5.1 Lessons from human motor control

The main problem in physical human-robot interaction is therefore tuning online the gains of the impedance or admittance control, for optimising performance, reducing the effort of the human operator⁴, and ensuring the stability.

Let us consider the simple case of the human arm interacting with a robotic manipulator at the end-effector. The main difficulty here for the robot is the accurate estimation of the varying environment, e.g., the end-point stiffness of the human arm [Ajoudani, 2016, Burdet et al., 2000]. To ensure stability and provide transparency, the impedance or admittance parameters should be tuned online according to the estimated human stiffness. A common way to estimate it, is via the grasping force at the interaction point, by dedicated force/torque sensors, even if in industrial applications there exists several cases where the human operator interacts directly with the payload, such as with lifting or assist devices. In [Lamy et al., 2009], the authors designed a handle for intuitive co-manipulation, where a 6-dof force/torque sensor was covered by a soft foam: when held delicately through a precision grip, the foam acts as a mechanical filter, and the system reflects low inertia; conversely, when held firmly through a power grasp, resulting from higher impedance in the arm because of muscle co-contraction [Napier, 1956], the squeezed foam does not filter, the system exhibits higher inertia, the gains are lowered and the system remains stable.

Human motor control actually provides several prior knowledge that can be exploited to adapt the impedance gains during physical interaction and more in general to optimize robot control [Ivaldi et al., 2012]. Dimeas and Aspragathos observed in [Dimeas and Aspragathos, 2016] that the bandwidth of voluntary motion in humans is relatively low and below 2Hz [de Vlught et al., 2003], hence during physical interaction with the robot it is possible to discriminate the human operator’s intent from the unstable motions thanks to frequency analysis. External oscillatory excitations of the arm up can be mitigated by the central nervous system (CNS) by changing the arm impedance: indeed, it has been shown by [E. Burdet, 2001] that the CNS can learn to control the magnitude, shape and orientation of the endpoint stiffness in a predictive way that is independent of the force needed to compensate for the imposed dynamics. Instead of simply co-contracting all the muscles, the CNS adopts the optimal strategy of increasing the impedance to be robust to external perturbations.

As discussed in [Yang et al., 2011], while in robots the common strategy to deal with perturbations and uncertainty is to design robust controllers that preserve stability, humans adapt the endpoint force and viscoelasticity to compensate for external forces, minimising error and effort while ensuring a constant stability margin. In fact, there is evidence that humans learn and adapt internal dynamic models of their own arm and of the environment during interaction [Shadmehr and Mussa-Ivaldi, 1994a], and these models are used for predicting muscle activations for planning movements. However, these feedforward signals alone are not sufficient in case of unstable dynamic environments: in several studies [Franklin et al., 2008, Franklin et al., 2007, Franklin et al., 2003], Franklin et al. reported evidence that the CNS regulates the arm compliance by co-contracting the muscles to cope with instability and feedback delays. More precisely, in [Franklin et al., 2007] the authors observed that the increase in stiffness was oriented along the direction of instability in the environment, confirming that the CNS is able to control the endpoint impedance of the limbs and selectively adapt it to the environment.

These studies support the idea that the CNS acts like an impedance controller at the level of the endpoint, ensuring stability, and reducing movement variability by increasing the impedance to reject disturbances. Upon these observations, in [Yang et al., 2011] the authors proposed a biomimetic controller with adaptive properties similar to those of humans. The controller increased feed-forward force and impedance in presence of large errors, whereas in absence of disturbance it decreased feed-forward force and impedance profiles while maintaining stability.

The aforementioned studies reported experimental evidence that the CNS can actually learn to control the interaction and adapt to different dynamics. In bimanual visuomotor tasks, even perturbed by deviating force fields, humans rapidly learn to control the interaction-forces by a combination of arm stiffness properties and direct force control [Squeri et al., 2010]. This property of the human mo-

⁴ The human effort can be identified in two stages: the physical and cognitive loads while learning a new collaborative task, and their amount on a regular basis after the human becomes an expert. This is important to note here that the CNS is capable of learning and adaptation to various tasks demands and disturbances, hence contributing to a reduction in the overall physical and cognitive loading while performing tasks with dynamic uncertainties [E. Burdet, 2001, Franklin et al., 2003]. Nevertheless, the (robot) counterpart’s adaptive behaviour can affect such learning and adaptation processes in terms of time and performance.
The question is very relevant as a continuous interaction will make the human “expert”, however this holds if the robot does not change its policy, forcing the human to re-adapt. There may be cases where a continuous re-adaptation from both sides would make sense and lead to more efficient collaboration, for example when the robot has to collaborate with a variety of different partners and could improve its skills by continuous learning from the different partners. However, to the authors’ knowledge there are not yet results showing the benefit of long-term multi-partner adaptation on collaborative tasks between robots and humans in real world applications.

6 Benchmarking and Relevant Use Cases

Over the last decade, several research groups aimed at evaluating the quality of human-robot interaction and collaboration by examining the acceptability of the framework by human volunteers. For example, authors in [Kahn et al., 2006] and [Feil-Seifer et al., 2007] proposed several benchmarking methods based on psychological assessments. In particular, a common approach, at the time, was to divide the assessment into three categories: one based on the robot (addressing safety and scalability), another based on social interaction (autonomy, imitation, privacy, understanding of specific domain, social success), and the latter based on the impact of the assistive technology on the users.

Other contributions additionally addressed the anthropomorphism of robotic agents, by confirming or invalidating the well-known concept of the Uncanny Valley [Hegel et al., 2008, Hwang et al., 2013]. The first studies were mainly focused on subjective responses, like questionnaires. Later on, based on the pioneering work of Kulic and Croft (see [Kulic and Croft, 2007]) the focus was moved towards a more objective evaluation, and no longer focused on pure appearance of robots, but rather on their motion capabilities (see also [Zanchettin et al., 2013a, Oztop et al., 2005, Kilner et al., 2003]). Recently, the authors in [Bestick et al., 2015] provided some objective insights on the benefits of a personalised robot behaviour on human physiological measurements. As of today, however, no significant evaluation procedures specifically targeting physical interaction is reported in literature.

When it comes to validating research results within effective and significant demonstrations, several benchmarking applications are introduced. Collaborative transportation of bulky and/or heavy objects is one the most common candidates to test collaboration modalities and interfaces. In most of the cases, the object to be manipulated is a rigid, non-articulated one, like a box or a planar surface, as reported in [Al-Jarrah and Zheng, 1997b, De Schutter et al., 2007, Wojtara et al., 2009, Mörtl et al., 2012, Agravante et al., 2014b]. The most challenging aspects of this demonstration are related to the task definition (both the human and the robot have to agree on the direction). Particularly interesting, from an ergonomic point of view, is on how the problem of sharing the carrying the effort is solved. For example in [Rozo et al., 2013] a learning strategy is combined with an impedance controller to perform a collaborative assembly task on a wooden table. In particular, while the robot is holding the table, the human can easily mount the legs as the robot automatically tunes its stiffness parameters to facilitate the completion of the task. The authors in [Kim et al., 2017] proposed a technique to monitor the overloading joint torque profiles (due to the external load) of the human partner for ergonomic and comfortable execution of heavy co-manipulation tasks. The authors of [Edsinger and Kemp, 2007] presented an object handover task between a robot and a human. The manipulator stiffness is adapted to accommodate for the physical interaction with the object and to ultimately accomplish the task.

Apart from the obvious domestic applications, this collaborative task has an immediate application in construction sites, as described in [Lee et al., 2007] where a mini-excavator was developed and used for the installation of curtain walls. Relevant applications of physical human-robot interaction are also reported in the healthcare field. Apart from rehabilitation robotics, which is not the target of this review paper, robotic assistance is the main relevant application. For example, in [Chuy et al., 2006] a collaborative mobile robot is developed in order to assist people to walk, despite physical disabilities. In [Ulrich and Borenstein, 2001] a haptic guidance cane is developed for the assistance of visually impaired people. A similar application scenario is also described in [Wakita et al., 2013].

In the field of industrial robotics, many manufacturing applications are also reported in literature, ranging from load-
ing and transporting heavy wheels [Levratti et al., 2016], collaborative assembly of a homokinetic mechanical joint [Cherubini et al., 2013, Cherubini et al., 2016] and cellular phones [Tan et al., 2009]. Another application within this field was presented in [Erden and Billard, 2014]. Here, an impedance controlled robot is used both to measure and learn the end-point impedance of expert welders for subsequent autonomous execution as well as for the training of non-skilled personnel. More recently, authors in [Peternel et al., 2016c, Peternel et al., 2016b] performed a collaborative sawing task as an illustrative example of dynamic industrial tasks with environmental uncertainties.

Some other developments are focused on collaborative aspects in manipulating non rigid or articulated objects [Colomé et al., 2015]. For example, in [Kosuge et al., 1998] the robot was commanded to deform a flexible metal sheet and support its payload so that the human could easily handle it. Similarly, in [Kruse et al., 2015] a collaborative robots was employed to help the human to fold a tablecloth. As for other examples, [Maeda et al., 2001] a collaborative rope turning was presented, while in [Donner and Buss, 2016a] the human and a robotic arm were collaboratively manipulating pendulum-like objects in order to establish a limit cycle with the desired amount of energy. Deformable materials are also of concern within domestic service robotics.

Finally, a recent trend in collaborative robotics research is devoted to the design of robots that take into account the ergonomics requirements typical of industrial applications [Maurice et al., 2017]. This field of research is not only promising but very important for the design of ergonomically compatible robots, as well as for the design of robot controls that optimize human ergonomics performances.

7 Discussions and outlook

The enhanced physical dexterity of the new generation of robotic platforms has paved the way towards their integration into robotics enabled service and care applications. Among various interaction scenarios, the exploitation of robots in collaborative settings has created a very important and high impact initiative. While HRC’s significant economic impact on industry is expected at large, it will also serve to maximise the social impact by maintaining the sense of purpose of the involved people in the work process.

In this paper, we provided an overview of the current state-of-the-art in human-robot collaboration. Although the hardware components of such systems are among the most crucial for a successful market entry, this review paper was dedicated to other important aspects, including the intermediate interfaces for improved human and robot perception; robot control modalities to perform on these channels (interaction modalities); the control performances (stability and transparency); and potential use cases (benchmarking).

The review of previous work on feed-forward (allowing the robot to perceive the human) and feed-back (complementing human perception of the task and the environment) communication channels indicated that the implementation of multi-modal human-robot interfaces presents a big advantage in addressing complex interaction scenarios. Nevertheless, it is likely to be accompanied by an increase on the software complexity and the human cognitive burden (for operating such modalities). Hence, an appropriate trade-off between the complexity and completeness of human sensorimotor behaviour and intention modelling must be defined. On the other hand, a considerable effort must be made to understand the required level of human awareness about the robot and environment states through artificial communication channels. In particular, despite the availability of several technologies, e.g., force feedback, augmented reality, etc., the amount of information (and its level of detail) the robot should communicate to the human is still an open research topic.

It was also indicated that the implementation of adaptive control methodologies such as impedance, force, admittance or hybrid approaches will not only enhance the robot adaptivity to the human and the environment, it may also lead to a reduced level of task-related pre-programming. Nevertheless, the application of such control concepts in human-robot collaboration still appears in a premature state and must be evaluated in cross-application HRC scenarios. In this direction, a principled human-robot mutual adaptation formalism is required to understand the levels of adaptivity from the two counterparts, and to address some questions arising on this topic: e.g., should the robot adaptation be generic or user-specific? As the robot eventually has some degree of controllability on the human companion, should this be still conceived as ethical?

Another very important aspect to be considered here is the underlying human and robot safety, so that well-performing robots can be successfully and seamlessly integrated in human-collaborative settings. Although this topic was not directly addressed in this review (due to the existence of a dense body if literature to discuss this aspect [Haddadin et al., 2009, De Santis et al., 2008, Alami et al., 2006]), it out to be mentioned that a great deal of effort must be directed towards ensuring safety for collaborating humans (to avoid injuries and accidents) and robots (to avoid unacceptable economic losses).

A list of potential applications and relevant use cases ranging from domestic to industrial environments was provided. A quick review of the state-of-the-art suggests that the research community’s attention is being extended from static tasks within structured environments to more dynamic collaborative scenarios that are subject to un-modelled dynamics and uncertainties, to facilitate a large-scale integration of robotic systems in several real-world scenarios.
Finally, several benchmarking examples were reported, with the underlying analysis mainly focusing on the acceptability aspects, either from an explicit (e.g. using questionnaires) or implicit (using physiological data) perspective. A natural problem arising from human-robot collaboration is to evaluate how the overall system behaviour, composed by the robot and the human, with respect to a certain set of performance indices. Such indices can be task-specific (a common approach in industry), however, this may limit the cross-application comparability of the HRC frameworks.

To conclude, this review paper was intended to give an updated overview of the state-of-the-art and recent research trends in human-robot collaboration. Despite the fast-growing interest in the real-world applications of the HRC, several challenges are still awaiting to be solved.

8 Acknowledgements

The authors would like to thank and remember Fabrizio Flacco for his spirit, contributions, and enthusiasm for writing this review paper. We will keep his memories alive in our hearts.

This work is supported in part by the EU FP7-ICT projects WALKMAN (no. 611832) and CoDyCo (no. 600716); in part by the H2020 Projects SoftPro (no. 688857) and AnDy (no.731540).

References


